

論文2002-39CI-2-6

# 실시간 정보 제공을 위한 웹 트랜스코딩

## (Web Transcoding for Real-Time Information Providing)

禹讚溢\*, 申仁澈\*\*

(Chan-Il Woo and In-Chul Shin)

## 요약

최근 들어 인터넷 접속을 위하여 웹 접근이 가능한 개인 장치들이 점점 더 이용 가능해 짐에 따라 PDA 나 셀룰러 폰과 같은 휴대 장치로부터 인터넷에 접속하는 사용자들이 증가하고 있다. 그러나 웹 상의 대부분의 페이지들은 적어도 640×480의 해상도를 가지는 컬러 모니터의 데스크탑 컴퓨터에 디스플레이하기 위해 제작되어, 충분한 표현 능력을 가지고 있지 않은 휴대 장치들에 동일한 내용의 웹 콘텐츠를 다르게 표현해야 할 필요가 요구되고 있다. 따라서 웹 콘텐츠와 휴대 장치들간의 이러한 문제를 해결하기 위하여 다양한 웹 트랜스코딩 기술이 제안되고 있다. 본 논문에서는 semi-automatic 웹 트랜스코딩 방법을 사용하여 실시간 정보를 위한 웹 트랜스코딩 시스템(웹 프리즘)을 제안한다. 본 논문에서 사용한 semi-automatic 트랜스코딩 방법은 모바일 사용자들이 웹 페이지의 레이아웃이 빈번하게 바뀌지 않는 제한된 범위의 웹 페이지에 접근하기 위한 좋은 방법으로 제공되며, 또한 웹 프리즘을 데스크탑 PC에서 구현하여 실험한 결과 PDA에서 좋은 품질의 문서를 생성하였다.

## Abstract

Recently, many web-enabled personal devices are becoming available for connecting to the internet, and there are many users who access the internet from hand-held devices such as PDAs, cell phones. Most pages on the world wide web are designed for displaying on desktop computers with color monitor having at least 640×480 resolution. But these devices do not have the same rendering capabilities as desktop computer, so the same web content needs to be rendered differently on hand-held devices. Therefore, there have been a variety of web transcoding techniques proposed to solve the problem between the web contents and hand-held devices. In this paper, we propose a web transcoding system(WebPrism) for real-time information providing using the semi-automatic transcoding method. The semi-automatic web transcoding method is a good solution when if mobile users access the limited number of web pages which layout is not changed frequently, also WebPrism is implemented as a desktop PC and experimental result shows that the proposed system generates the best looking document for a given display size of PDA.

Key Words : 웹트랜스 코딩, 무선 인터넷

\* 正會員, 檀國大學校 大學院 電子工學科  
(Department of Electronic Engineering, Graduate School, Dankook University)

\*\* 正會員, 檀國大學校 電氣電子컴퓨터工學科  
(Department of Electrical, Electronic and Computer Engineering, Dankook University)

接受日字:2001年12月3日, 수정완료일:2002年2月21日

## I. 서론

미국의 ARPANET으로부터 출발한 인터넷은 월드 와이드 웹(WWW)이 등장한 이래 웹 사이트와 그 사용자의 수가 매년 폭발적으로 증가하여 왔으며 인터넷 초기부터 지금까지 인터넷에 접속할 수 있는 디바이스는 컴퓨터가 거의 유일하였다. 그러나 최근 들어 PDA

(personal digital assistant)나 셀룰러 폰과 같은 무선 단말기들이 인터넷에 접속할 수 있는 기능을 갖추므로써 데스크탑 PC를 대상으로 작성된 웹 페이지들을 무선 단말기에서도 표현해야 할 필요성이 요구되고 있다.<sup>[1]</sup> 그러나 대부분의 웹 페이지들은 뛰어난 처리 능력을 가지고 있는 데스크탑 PC를 기반으로 하여 적재는 640×480 해상도에서 1024×768의 고 해상도를 지원하는 15인치 이상의 모니터를 대상으로 구성되어 있어 상대적으로 낮은 처리 능력과 작은 크기의 디스플레이를 가지는 무선 단말기에서 기존의 웹 페이지를 표현하기에는 상당한 제약이 따른다.<sup>[2,3]</sup> 따라서 현재 운영되고 있는 웹 페이지들을 제대로 보여주지 못하는 문제점이 있어 PDA나 셀룰러 폰 등의 무선 단말기들은 상대적으로 낮은 버전의 HTML만을 지원하거나 HTML을 변형한 Markup Language를 지원하고 있어 각각의 휴대 장치에 맞는 언어로 변형해야 한다.<sup>[4,5]</sup>

국내의 경우 SK 텔레콤(011, 017)에서는 XML을 기반으로 하여 만들어진 WML (wireless markup language)을 사용하고 있으며, KTF(016, 018)의 경우 mHTML(mobile HTML)을 사용한다. 따라서 많은 웹 사이트들이 필요에 따라 이러한 무선 단말기들을 위한 전용 웹사이트를 구축하고 있다. 그러나 전용 웹사이트를 새로 구축하는 것은 비용 적인 면이나 시간적인 면에서 매우 비효율적일 뿐만 아니라 특히 앞으로 계속 등장할 새로운 디바이스들에 일일이 대응해야 한다는 점을 고려하면 바람직하지 못하다. 따라서, 이러한 문제점들을 해결하기 위하여 웹 트랜스코딩(transcoding) 기술이 널리 연구되고 있다.<sup>[6-8]</sup>

웹 트랜스코딩 기술은 클라이언트 측(client-side) 기술과 서버 측(server-side) 기술로 나누어지며 클라이언트 기술은 휴대 장치 내에서 웹 페이지를 변환하는 것으로 휴대 장치의 제한된 성능으로 인하여 zooming, table unrolling 등의 제한된 범위의 트랜스코딩 기술이 적용된다. 그리고 서버 기술은 HTTP proxy 서버 상에서 웹 페이지를 휴대 장치의 디스플레이 크기에 맞추어 자동으로 변환하는 것으로 여러 가지 트랜스코딩 기술이 적용된다. 현재 무선 단말기를 이용하여 웹 접속을 위한 서비스는 무선 단말기의 제한된 성능으로 인하여 매우 제한된 범위 내에서 이루어지고 있으며, 무선 단말기를 사용하여 인터넷에 접속하는 사용자들은 웹 페이지의 전체 내용보다는 관심 있는 특정 내용만을 보고자 하는 것이 일반적이다. 따라서 본 논문에서

서는 이러한 관점에서 240×320의 해상도를 가지는 PDA를 대상으로 기존의 웹 페이지들 중에서 실시간(real-time)으로 변하는 정보(뉴스, 교통정보)에 대한 semi-automatic 웹 트랜스코딩 방법을 제안한다.

## II. 웹 트랜스코딩

### 1. 웹 트랜스코딩 기술.

트랜스코딩이란 특정 형식의 데이터를 다른 형식으로 변환해 주는 것을 말하며, 웹 트랜스코딩은 PC를 대상으로 만들어진 기존의 웹 콘텐츠들을 무선 단말기 특성에 맞추어 표현할 수 있도록 변환해 주는 기술을 말한다. 웹 트랜스코딩은 웹 페이지의 내용을 휴대 장치 크기에 맞추어 자동으로 분할 및 축소하는 과정으로 텍스트나 이미지의 제목 등은 사용자에게 보여주고 분할된 내용들은 하이퍼링크(hyperlink)로 처리하여 사용자에게 의한 디스플레이 요구가 있을 경우 무선 단말기 화면에 표시한다.

웹 콘텐츠를 PDA에서 표현 하고자 할 경우 다음과 같은 문제점들이 발생한다.

첫째, 기존의 웹 페이지들은 640×480 이상의 해상도로 작성되어 있어 240×320의 해상도를 가지는 PDA 화면에 맞추어 표현할 수 없다.

둘째, 웹 페이지들은 높은 처리 능력을 갖는 데스크탑 PC를 기반으로 작성되어 있어 상대적으로 낮은 처리 능력을 갖는 PDA는 웹 페이지를 읽어와서 디스플레이 하기 위한 처리 속도가 떨어진다.

셋째, 웹 페이지 내에 포함된 이미지는 크기(size)와 용량이 다양하다. 따라서, PDA에서 이미지를 읽어올 경우 화면의 크기와 맞지 않을 경우가 발생하며 이미지 데이터의 로딩(loading) 속도가 월등히 떨어진다.

현재 PDA에서 웹 콘텐츠의 내용을 볼 수 있는 방법은 두 가지로 나뉘어진다. 그 중 하나는 PDA에 무선 모뎀을 장착하여 실시간으로 웹 페이지에 접근하는 방법이다. 이 방법은 비싼 무선 통신 사용료를 지불해야 되고 PDA 전용 웹 페이지나 PDA의 디스플레이에 맞추어 트랜스코딩 된 페이지가 아닌 일반 웹 페이지를 다운 받았을 경우 디스플레이 하는데 많은 처리 시간이 소요된다. 그리고 최근 들어 PDA를 사용하여 무선으로 인터넷에 접근하기 위한 방법으로 블루투스(bluetooth)를 내장한 제품들이 출시되고 있어 별도의

통신비용 없이 웹 페이지에 접근할 수 있는 방법을 제공하며 기존의 블루투스에서 지원되지 않던 Handoff/Roaming 기술을 사용하여 거리제한을 해결할 수 있는 방법이 개발되어 블루투스가 탑재된 휴대용 기기를 무선 랜과 같이 사용할 수 있다.<sup>[9,10]</sup> PDA를 통한 웹 접근의 또 다른 한 가지 방법은 아방고(avantgo)<sup>[11]</sup>나 아이핸디고(ihandygo)<sup>[12]</sup>와 같은 서비스를 이용하는 방법이 있다. 이 방법은 PDA를 PC와 유선으로 연결한 후 서비스 업체에서 제공하는 정보를 PDA로 다운 받아 저장한 후 분리 휴대하여 이동하며 저장된 내용을 볼 수 있는 방법이다.

웹 트랜스코딩 기술은 서버 측 기술과 클라이언트 측 기술의 두 개의 범주로 나뉘어진다. 클라이언트에서 수행되는 웹 트랜스코딩 기술은 휴대장치들이 HTTP 서버로부터 HTML 페이지의 전체 내용을 수신하고, 휴대장치 내에서 콘텐츠(contents)의 내용을 변형시킨다. 반면에 서버에서 수행되는 웹 트랜스코딩 기술은 HTTP 서버 내에서 웹 콘텐츠를 휴대 장치 크기에 맞게 변형시켜 휴대장치들은 축소되고, 개선된 웹 콘텐츠를 수신 받을 수 있도록 하는 기술로 manual, semi-automatic, fully-automatic 방법으로 나누어진다.<sup>[13,14]</sup>

#### (1) 클라이언트 기술.

클라이언트에서 수행되는 웹 트랜스코딩 기술은 제한된 성능을 가진 휴대 장치 내에서 트랜스코딩이 수행되며 트랜스코딩의 여러 가지 기술 중 제한된 범위의 트랜스코딩 기술이 사용되어 트랜스코딩 된 페이지의 품질이 나빠진다. 클라이언트에서 수행되는 트랜스코딩 기술은 zooming, table unrolling<sup>[15]</sup> 등을 들 수 있으며 웹 페이지 구조가 간단할 경우 유용하게 사용될 수 있는 방법이지만 웹 페이지의 구조가 복잡한 형태일 경우 사용할 수 없다.

#### (2) 서버 기술.

클라이언트에서 수행하는 웹 트랜스코딩 기술은 휴대 장치들이 가지는 제한된 성능으로 인하여 제한된 범위 내에서 트랜스코딩이 수행되지만 서버에서 수행되는 웹 트랜스코딩은 그러한 제한을 갖지 않으며 다음과 같은 방법으로 나누어진다.

##### ① Manual 방법.

이 방법은 웹 페이지의 내용을 특정 장치에 맞게 새로 제작하는 것으로, 각 장치의 특성을 고려하기 때문에 높은 품질의 트랜스코딩 된 페이지를 얻을 수 있다.

그러나 새로 제작된 페이지만을 접근할 수 있는 단점이 있으며, 또한 각 장치의 특성이 변하거나 웹 페이지가 업데이트(update) 되면 트랜스코딩 된 페이지는 새로 만들어져야 하기 때문에 매우 불편하다.

##### ② Semi-automatic 방법.

일반적으로 대부분의 웹 페이지들은 특정한 키워드(keyword)나 규칙적인 표현들을 사용하여 웹 콘텐츠를 제작한다. 따라서, 페이지 필터링(page filtering)이라고도 호칭되는 semi-automatic 방법은 웹 페이지내의 주석, 특정 키워드 또는 규칙적인 표현을 사용하여 웹 페이지를 트랜스코딩 한다. 이 방법은 모바일 사용자가 웹 페이지 전체 레이아웃(layout)이 빈번하게 변하지 않는 제한된 웹 페이지를 접근할 때 좋은 방법을 제공하고 있으며 높은 품질의 트랜스코딩 된 페이지를 얻을 수 있다. 그러나 웹페이지의 레이아웃(layout)이 자주 바뀔 경우 manual 방법과 같은 단점이 있다.

##### ③ Fully-automatic 방법.

웹 페이지를 자동으로 변환시켜 주는 방법으로 휴대 장치로부터 모든 웹 페이지를 접근할 수 있어 manual과 semi-automatic 방법의 문제점을 해결할 수 있다. 그러나 manual과 semi-automatic 방법과 비교하면 트랜스코딩 된 페이지의 품질이 나빠지며, 현재 개발된 fully-automatic 트랜스코딩 기술은 흔히 매우 나쁜 품질의 트랜스코딩 된 페이지를 생성하기도 한다.

#### 2. 아방고 서비스<sup>[11]</sup>

아방고 서비스는 무선 모뎀 없이도 PDA에서 웹 페이지의 내용을 볼 수 있도록 하는 웹 브라우징 서비스로, 아방고 이용자는 PDA를 PC와 연결하여 자신이 선택한 웹사이트(PDA에서 볼 수 있도록 재구성되어 있는 사이트)를 PDA로 다운 로드 받아 볼 수 있으며, PDA를 PC와 연결시킬 때마다 새로 업데이트 된 웹사이트의 내용이 PDA로 다운로드 된다. 아방고와 같은 서비스는 사용자들에게 양질의 콘텐츠를 제공하여 줄 수 있지만 웹 페이지의 내용이 변하게 되면 PC와 연결하여 다시 웹 페이지를 다운 받아야 되고 또한, PDA의 제한된 메모리 용량으로 인하여 웹 페이지의 일부만 저장할 수밖에 없는 단점이 있다.

아방고 서비스는 다음과 같이 운용된다.

- ① 아방고 홈페이지에 접속한 후 사용자 계정 생성.
- ② 아방고 프로그램 설치.

- ③ 사용자가 원하는 웹사이트를 아방고 채널로 등록.
- ④ PDA를 PC와 연결한 후 아방고 사이트로부터 웹 페이지 다운로드.

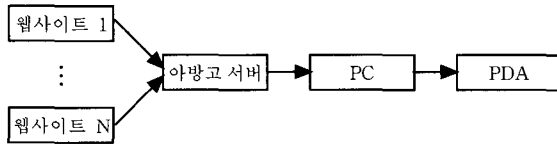


그림 1. 아방고 서비스  
Fig. 1. Avantgo services.

### III. 웹 트랜스코딩 서버 기술

웹 콘텐츠를 사용자의 휴대 장치 크기에 맞추어 변환하는 기본적인 웹 트랜스코딩 기술은 다음과 같은 과정을 수행한다.<sup>[13,14]</sup>

#### 1. Outlining transform.

Outlining transform은 신문 기사와 같이 문단 헤더(header)가 포함된 문단을 트랜스코딩 하는데 매우 효과적으로 사용되는 기술이다. 이 방법이 적용된 웹 페이지는 각 문단에서 헤더만 남겨두고 문단의 내용은 제거된다. 그리고 제거된 문단을 사용자들에게 보여주기 위하여 각 헤더를 하이퍼링크(hyperlink)로 변환하여 사용자가 헤더를 선택하였을 경우 링크 된 각 문단의 내용을 보여준다.

그림 2는 outlining transform을 나타내고 있다. 여기서, 그림 (a)는 원래의 웹 페이지의 내용을 보여주고 있으며 그림 (b)는 각 문단에서 문단 내용이 제거된 헤

더만을 PDA에서 보여주고 있는 것을 나타내며, 제거된 각 문단의 내용은 문단의 헤더(Section 1, Section 2)를 선택하였을 경우 PDA에 디스플레이 된다.

#### 2. First sentence elision transform.

대부분의 웹 페이지들은 텍스트 블록들을 포함하고 있으나 텍스트 블록의 각 문단은 헤더를 포함하지 않을 경우가 있다. 따라서 이러한 텍스트 블록들을 특정 디스플레이 크기로 축소하기 위하여 first sentence elision transform을 사용한다. 이 방법은 outlining transform과 비슷하나 텍스트 블록에 헤더가 없어 각 텍스트 블록의 첫 문장을 하이퍼링크로 변환하여 사용자에게 디스플레이하고 나머지 문단은 제거한다. 그리고 제거된 문단은 사용자가 디스플레이 되는 첫 문장을 선택하면 하이퍼링크에 의하여 디스플레이 된다.

#### 3. Indexed segmentation transform.

Indexed segmentation transform은 휴대 장치별로 설정된 사이즈를 초과하는 콘텐츠를 여러 개의 서브 페이지로 나누어 첫 번째 서브 페이지의 콘텐츠만을 서비스하고 각 서브 페이지별로 링크를 달아 사용자가 차례로 볼 수 있도록 한다.

Indexed segmentation transform에서는 먼저 현재 입력 페이지가 휴대 장치에 표시되기에 충분히 작은 크기인지를 체크한 다음, 휴대 장치보다 크다면 text blocks, lists와 같은 로직컬(logical)한 요소들을 찾은 후, 휴대 장치의 디스플레이 크기에 맞도록 로직컬 한 요소들을 1:1로 매칭 하여 새로운 페이지를 생성한다. 만약 분할된 하나의 로직컬 한 요소들이 휴대 장치의

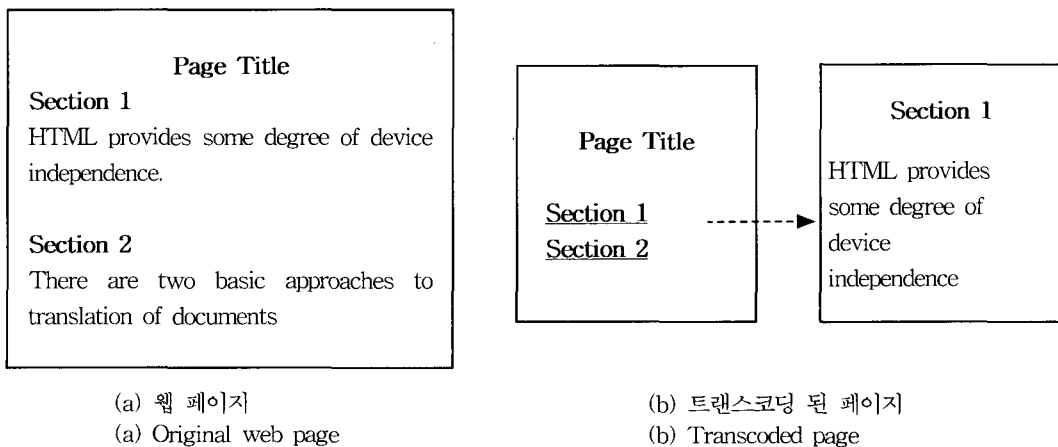


그림 2. 아웃라인 변환  
Fig. 2. Outlining transform.

디스플레이에 맞지 않으면 추가적으로 분할(segmentation) 과정이 수행된다. 이 경우, 문단이나 테이블 구조 등은 파괴될 수 있다.

4. Table transform.

Table transform은 웹 문서의 테이블 구조에 대하여 다루는 것으로 먼저, 웹 페이지 내에서 테이블이 위치한 곳을 찾은 후 테이블이 어떠한 변환 없이 전송될 수 있는가를 검사한다. 만약 디스플레이를 위하여 변환 과정이 필요하면 테이블 셀(cell) 당 하나의 서브 페이지를 생성하는데 각 셀은 top-down and left-to-right 순서로 진행하여 하이퍼링크로 연결한다. 이 방법에서는 테이블 내의 셀 들이 분리되어 표현되기 때문에 셀 들 간의 상호 관계를 이해하는 것이 거의 불가능해진다.

5. Image reduction and elision transforms.

웹 문서의 이미지를 다루는데 유용한 기술로, 크기가 큰 이미지는 미리 정의된 스케일 값(scaling factor)으로 축소시키고 축소된 이미지는 하이퍼링크로 연결한다. 만약 이미지를 축소한 후에도 휴대 장치의 디스플레이 크기에 적당하지 않으면 다음의 3가지 제거 방법이 사용된다.

- Elide All policy.

모든 이미지들을 하이퍼링크로 대체.

- First Image Only policy

첫 번째 이미지를 제외한 모든 이미지들을 하이퍼링크로 대체.

- Bookends policy

첫 번째와 마지막 이미지들을 제외하고 모든 이미지를 하이퍼링크로 대체.

6. Image map transform.

휴대 장치에서 이미지를 표시할 수 없는 경우 웹 문서로부터 이미지를 제거하고 웹 문서의 이미지 링크 부분에서 ALT 태그(tag)를 검사한 후 ALT 태그의 텍스트를 추출하여 이미지 대신 디스플레이 한다. HTML

의 ALT 태그는 이미지가 전송되지 않았을 경우를 대비하여 이미지 대신 태그 안의 텍스트를 디스플레이 하기 위해 다음과 같이 사용된다.

```
<IMG SRC="이미지 경로" ALT="텍스트">
```

IV. 제안된 웹 트랜스코딩 기술

기존에 제안된 웹 트랜스코딩 방법은 트랜스코딩 시스템이 HTTP Proxy 서버에 위치하여 사용자의 요구가 있을 경우 해당 URL의 웹 문서와 이미지 등을 가져와서 프록시 서버 상에서 웹 페이지를 휴대 장치에 맞게 변환하고 변환된 문서나 이미지를 사용자의 휴대 장치로 전송한다. 이러한 방법은 모든 웹사이트의 페이지들을 휴대 장치 크기에 맞게 표현할 수 있는 장점이 있으나 새로운 페이지에 대한 트랜스코딩 요구가 들어오면 트랜스코딩 하는데 시간이 소요되며, 트랜스코딩 된 문서의 품질 또한 많이 떨어지게 되는 단점이 있다.

본 논문에서는 휴대 장치를 이용하여 인터넷에 접속하는 사용자들은 PC 환경에서처럼 웹 페이지의 전체 내용보다는 페이지 내의 특정 정보를 얻기 원한다는 점에 착안하여 뉴스 속보, 교통 정보와 같이 실시간으로 변하는 정보들을 제공하는 웹 페이지들을 대상으로 semi-automatic 웹 트랜스코딩 기술을 적용하여 페이지 내의 일부 내용만을 트랜스코딩 하여 PDA 사용자들이 실시간으로 변하는 속보를 검색할 수 있는 웹 트랜스코딩 기술(본 논문에서 웹 프리즘(WebPrism)으로 명명)을 제안한다. 본 논문에서 제안하는 웹 프리즘 시스템은 그림 3과 같이 구성되어 있다.

웹 프리즘은 실시간 속보를 제공하는 URL로부터 속보를 PC로 다운 받은 후 트랜스코딩 과정을 수행하여 변환된 정보를 서비스할 수 있는 웹 서버로 전송한다. 따라서 PDA 사용자들은 웹 서버에 접속하여 실시간 속보를 전송 받을 수 있다. 웹 프리즘은 환경 설정, 웹 페이지 다운로드, 웹 트랜스코딩, FTP 전송 등의 모듈

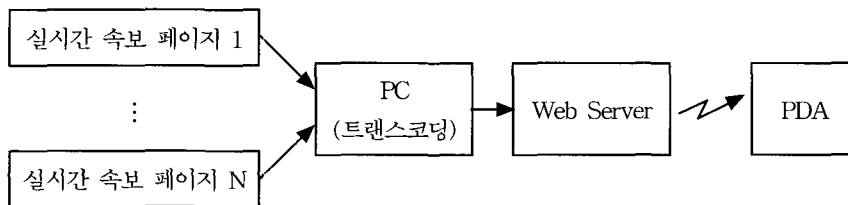


그림 3. 웹 프리즘 시스템  
Fig. 3. WebPrism system.

로 구성되어 있다.

### 1. 환경 설정.

환경 설정 부분은 웹 프리즘 시스템의 전체 모듈 동작에 필요한 사항들을 설정하는 부분으로 다음과 같이 구성한다.

- 실시간 속보 페이지의 URL 입력.
- 실시간 속보 구분을 위한 정보 입력.

웹 페이지 내에서 실시간 속보가 표현되는 부분을 찾기 위하여 <LI>나 <UL>과 같은 문장 목록을 표시하는 태그를 체크하는데 페이지에 따라서 이러한 태그 대신 주석이나 다른 키워드를 사용할 수 있다. 따라서 트랜스코딩 모듈에서 실시간 속보가 있는 부분을 찾기 위한 키워드를 입력받는다.

- 실시간 속보의 다운로드 시간 입력.

실시간 속보는 몇 초 또는 몇 분 간격으로 계속해서 변하므로 변환된 정보를 일정한 간격으로 다운로드 하기 위한 시간을 설정한다.

- 이미지 데이터의 변환 여부 설정.

실시간 속보를 제공하는 웹사이트는 일반적으로 첫 페이지에서는 시간대별로 제공되는 속보 제목을 나타낸다. 그리고 속보 제목을 클릭(click)하면 해당 기사가 있는 웹 페이지로 연결된다. 속보 내용은 일반적으로 텍스트로 구성되어 있으나 이미지 텍스트로 표현되는 경우도 있다. 이러한 이미지 기사는 PDA의 디스플레이 크기로 축소하였을 경우 알아 볼 수 없을 정도로 이미지 내의 글자가 축소 될 수 있다. 따라서 이러한 경우 이미지 데이터를 디스플레이 하는 것은 의미가 없을 수 있으며 또한, 이미지 데이터는 수십KB 이상의 용량으로 구성되어 있어 PDA에서 다운로드 하여 디스플레이 하는데 많은 시간이 소모된다. 따라서, 이러한 정보의 디스플레이 여부를 설정한다.

- FTP 전송을 위한 웹 서버 주소 및 서버 ID, 패스워드 입력.

트랜스코딩 된 정보를 웹 서버로 전송하기 위하여 웹 서버 주소와 ID, 패스워드를 입력 받는다.

### 2. 웹 페이지 다운로드.

트랜스코딩을 위하여 실시간 속보가 있는 페이지를 해당 URL로부터 다운 받는다.

### 3. 웹 트랜스코딩.

웹 트랜스코딩 모듈은 앞에서 설명한 트랜스코딩의 기본 방법들 중에서 outlining transform, image

reduction and elision transforms, image map transform, table transform을 사용하며 first sentence elision transform과 indexed segmentation transform은 사용하지 않는다. First sentence elision transform은 링크되는 속보 내용을 디스플레이 하는데 사용할 수 있는데 셀룰러 폰과 같이 작은 크기의 디스플레이를 가지는 휴대 장치에 표시하는데는 유용할 수 있지만 240×320 크기를 가지는 PDA에서는 스크롤 바를 이용하여 검색하는데 큰 불편함이 없으며 만약, 속보 내용을 여러 개의 작은 페이지로 나누어 디스플레이 할 경우 사용자 입장에서는 속보 내용이 끊겨서 보여지게 되므로 불편함이 따른다. 따라서 이 방법은 본 논문에서는 사용하지 않는다. 그리고 indexed segmentation transform의 경우는 웹 페이지를 작은 크기의 서브 페이지들로 분할하기 위하여 사용되는데 본 논문에서는 웹 페이지의 전체 내용보다는 속보와 관련된 정보만 선택하기 때문에 사용하지 않는다.

#### ① Outlining transform.

기존의 outlining transform은 각 문단을 제거하고 문단 제목을 하이퍼링크로 변환하여 디스플레이 한 후 사용자의 선택이 있을 경우 제거된 문단을 디스플레이 하기 위하여 사용한다. 하지만 본 논문에서는 실시간 속보 제목과 내용이 서로 다른 페이지로 구성되어 있어서 속보 제목을 찾기 위하여 outlining transform을 사용하며, 속보 내용은 속보 제목의 하이퍼링크를 검사하여 읽어들인다. 여기서, 속보 제목과 연결된 하이퍼링크가 상대 경로를 사용할 경우 다음과 같이 절대 경로를 붙여준다.

```
<a href="/news/2001/11/02/all/20011102161021102321.html">
<a href="www.joins.com/news/2001/11/02/all/20011102161021102321.html">
```

#### ② Image reduction and elision transforms.

이 방법은 속보 내용을 디스플레이 하기 위한 방법으로 사용한다. 만약 속보 내용이 이미지 기사일 경우 이미지의 크기를 PDA의 디스플레이(240×320) 크기와 비례하여 축소한다. 그러나 이미지 기사가 필요 없을 경우 관련된 이미지 기사는 제거된다.

#### ③ Image map transform.

하이퍼링크로 연결되는 정보가 이미지 일 경우 이미지가 표시되지 않을 수 있다. 이러한 경우 ALT 태그의 텍스트 부분을 추출하여 이미지 대신 출력한다.

#### ④ Table transform.

Table transform은 테이블 내에서 속보 제목과 속보 내용이 있는 셀을 검출하기 위하여 사용한다. 일반적으로 속보 제목과 속보 내용은 테이블 내의 셀 중에서 가장 넓은 폭(width)을 차지한다. 따라서, 테이블 내의 각 셀을 검사하여 가장 넓은 셀을 검색하고, 입력된 키워드를 사용하여 속보 제목과 내용을 검사한다.

#### 4. FTP 전송.

실시간 정보가 있는 URL로부터 다운 받은 웹 페이지는 트랜스코딩 과정을 수행한 후 PDA 사용자들에게 서비스하기 위하여 웹 서버로 전송된다. 따라서 PDA 사용자들은 트랜스코딩 된 다양한 실시간 정보가 있는 웹 서버로 접속하여 작은 크기의 트랜스코딩 된 페이지를 PDA에 디스플레이 한다.

### V. 실험 및 결과

본 논문에서 제안한 웹 프리즘은 Pentium III(1GHz)에서 JAVA로 구현하였으며, 주요 클래스 및 메서드의 구성은 다음과 같다.

#### ① WebPrism.class

웹 프리즘의 메인 모듈로서 전체화면을 구성하고 PlayClock 클래스의 객체를 생성 하여 현재 시간을 체크한 후 GetURL 클래스의 url() 메서드를 호출하여 설정된 URL의 HTML 페이지를 읽어온다. 그리고 트랜스코딩 과정을 수행한 후 트랜스코딩 된 페이지를 웹 서버로 전송하기 위하여 Ftp() 메서드를 호출하는 모듈들로 구성된다.

##### 1) fullDownMenu()

웹 프리즘의 메뉴를 구성하며 각 메뉴에 대한 이벤트 발생시 이벤트 핸들러(event handler)에게 이벤트를 전달한다.

##### 2) connectURL()

GetURL 클래스의 객체를 생성하고, GetURL 클래스의 생성자에 웹 페이지의 URL을 인자로 넘겨준다.

##### 3) Ftp()

트랜스코딩 된 페이지를 웹 서버로 전송한다.

##### 4) initTimeThread()

Runnable 인터페이스를 구현한 쓰레드 클래스인 PlayClock 클래스의 객체를 생성하여 Thread 클래스 생성자의 인자로 주어 Thread를 생성하고, 이

Thread 클래스 객체의 start() 메서드를 호출한다.

#### ② PlayClock.class

현재 시간을 체크한 후 메인 프레임에 디스플레이하고 설정된 URL의 페이지를 읽어오기 위한 시간을 체크하여 WebPrism 클래스의 connectURL() 메서드를 호출한다.

#### ③ GetURL.class

지정된 URL로 부터 HTML 페이지를 읽어오는 모듈.

##### 1) url()

GetURL 클래스 생성자의 인자로 넘어온 URL 값을 URL 클래스 생성자의 인자로 주어 해당 웹 페이지를 읽어온다.

#### ④ Converter.class

트랜스코딩 수행을 담당하는 클래스로서 페이지내의 HTML을 읽어와서 불필요한 태그들을 제거한 후 실시간 기사와 관련 링크 등의 데이터를 추출하여 트랜스코딩 된 페이지를 생성한다.

##### 1) transcoding()

의미 없는 정보와 불필요한 태그를 삭제하고 실시간 정보와 관련 링크 등의 데이터를 추출하여 트랜스코딩 된 페이지를 생성한다.

##### 2) updateFile()

트랜스코딩 된 웹 페이지의 HTML 코드 상, 하단 부분에 특정 로고나 기타 정보를 PDA에서 디스플레이하기 위하여 삽입한다.

실험에 사용된 실시간 정보를 제공하는 웹사이트를 표 1에 보였다.

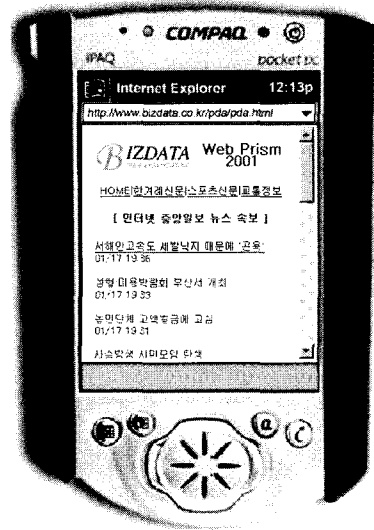
표 1. 실험에 사용된 웹 페이지  
Table 1. Web pages tested in this paper.

중앙일보 속보	http://www.joins.com/ml/ml.html
한겨레신문 속보	http://www.hani.co.kr/TOTAL/home01.html
한국일보 속보	http://www.hankooki.com/ds_listall/page_quickports.htm
고속도로 교통정보	http://www.freeway.co.kr

그림 4는 PC 모니터 상에서 표현되는 중앙일보 속보 페이지와 PDA에서 표현되는 트랜스코딩 된 페이지를 나타내고 있다. 실험 결과 그림에서 나타나는 바와 같이 PDA 용으로 변환된 웹 페이지의 품질은 우수하게 나타남을 알 수 있다.



(a) 중앙일보 실시간 속보 페이지  
(a) Joong-ang il bo real-time news page



(b) PDA 화면  
(b) Display of PDA  
그림 4. 실시간 속보 페이지의 예.  
Fig. 4. An example of the real-time information page.

PDA 사용자들을 위한 오프라인 웹 브라우징 서비스인 기존의 아방고 서비스는 아방고 서비스를 지원하는 인터넷 사이트의 웹 페이지(PDA의 디스플레이 크기에 맞추어 변환된 웹 페이지)를 온라인 상으로 제공하며,

PDA 사용자들은 PC와 PDA를 연결한 후 변환된 페이지를 PDA로 다운 받아 저장한 후 이동하면서 아방고 전용 프로그램을 사용하여 저장된 내용을 볼 수 있는 방법을 제공한다. 그러나 이 방법은 웹 페이지에 대한 새로운 내용의 서비스를 받기 위해서는 항상 PC와 연결해야 되기 때문에 실시간 정보에 대한 서비스가 불가능하며 PDA의 저장 용량의 한계로 제한된 수의 웹 페이지만 저장되는 단점이 있다.

그러나 본 논문에서 제안한 웹 프리즘은 웹사이트에서 제공하는 실시간 정보를 주기적으로 다운 받아 웹 트랜스코딩 과정을 수행한 후 트랜스코딩 된 정보를 웹 서버로 전송하여 PDA 사용자들에게 서비스하기 때문에 무선으로 웹 접속이 가능한 PDA 사용자들은 트랜스코딩 된 웹 페이지를 별도로 저장할 필요 없이 PDA용 웹 브라우저를 사용하여 언제든지 실시간 정보를 볼 수 있는 장점이 있다.

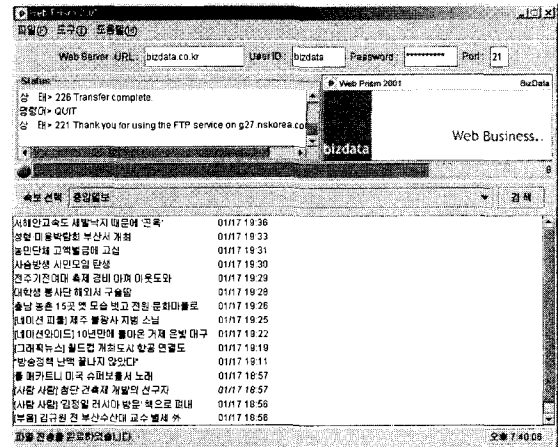


그림 5. 웹 프리즘  
Fig. 5. WebPrism.

그림 5는 본 논문에서 제안한 웹 프리즘이 동작하는 과정을 보여주고 있다. 그림 5에서 상단의 텍스트 입력 부분은 트랜스코딩 된 페이지를 웹 서버로 전송하기 위하여 웹서버 주소와 사용자 ID, 패스워드를 입력받는 부분과 FTP 접속을 위한 포트 번호를 나타낸다. 또한 "Status"로 표시된 텍스트 영역(text area)은 트랜스코딩 된 웹 페이지를 웹 서버로 전송하는 상태와 웹 서버로부터 전달받은 상태를 나타낸다. 그리고 중앙의 진행바(scroll bar)는 지정된 URL로부터 파일을 다운 받아 트랜스코딩을 수행하기 위한 시간간격을 나타내며 우측의 숫자는 트랜스코딩 과정이 수행된 횟수를 나타



낸다. 마지막으로 트랜스코딩 된 페이지를 웹 프리즘에서 읽어들이 디스플레이 하는 것을 그림 5의 하단부 텍스트 영역에서 나타내고 있다.

## VI. 결 론

PDA나 셀룰러 폰과 같은 무선 단말기들이 인터넷에 접속할 수 있는 기능을 갖추으로써 데스크탑 PC를 대상으로 작성된 웹 페이지들을 무선 단말기에서도 표현해야 할 필요성이 요구되고 있다. 그러나 휴대 장치들은 디스플레이 크기와 처리 능력이 데스크탑 PC와 달라 제한된 성능을 가진 휴대 장치들에 기존의 웹 페이지를 표현하는데 문제가 발생한다. 따라서, 이러한 문제점들을 해결하기 위하여 웹 트랜스코딩 기술이 제안되었다. 웹 트랜스코딩 기술은 클라이언트 기술과 서버 기술로 나누어지며 클라이언트에서 수행되는 웹 트랜스코딩 기술은 웹 트랜스코딩의 여러 가지 기술 중 제한된 범위의 트랜스코딩 기술이 사용되어 트랜스코딩된 페이지의 품질이 나빠지는 단점이 있으며, 서버에서 수행되는 웹 트랜스코딩 기술 중 fully-automatic 방법은 모든 웹 페이지들을 휴대 장치 크기에 맞게 표현할 수 있는 장점이 있으나 트랜스코딩 된 문서의 품질이 많이 떨어지는 단점이 있다.

본 논문에서는 뉴스 속보, 교통 정보와 같이 실시간으로 변하는 정보들을 제공하는 웹 페이지들을 대상으로 semi-automatic 트랜스코딩 기술을 적용하여 페이지 내의 일부 내용만을 트랜스코딩 하여 PDA 사용자들이 실시간으로 변하는 정보를 검색할 수 있는 웹 트랜스코딩 방법을 제안하였다. PDA 사용자를 위한 아방고 서비스는 서비스 업체에서 제공하는 정보를 PDA로 다운 받아 저장한 후 분리 휴대하여 이동하며 저장된 내용을 볼 수 있는 방법을 제공한다. 그러나 이 방법은 새로 업데이트 된 내용을 보기 위하여 PDA를 PC와 연결한 후 새로운 페이지를 다운 받아야하는 단점이 있다. 그러나 본 논문에서 제안한 방법은 이러한 단점을 개선하여 실시간으로 정보를 제공할 수 있는 방법을 제공한다.

본 논문에서 사용한 semi-automatic 트랜스코딩 방법은 웹 페이지 내에서 사용한 규칙적인 표현이나, 특정 키워드 등을 검사하여 트랜스코딩을 수행하는 것으로 웹 페이지의 전체 레이아웃(layout)이 빈번하게 변하지 않는 제한된 범위의 웹 페이지를 접근할 때 좋은

방법을 제공하고 있으며 높은 품질의 트랜스코딩 된 페이지를 얻을 수 있다. 실험 결과 본 논문에서 제안한 웹 프리즘은 매우 우수한 품질의 트랜스코딩 된 페이지를 생성하여 PDA에서 표현하는데 아무런 문제점이 없음을 확인하였으며, 실시간으로 변하는 정보를 주기적으로 업데이트하기 때문에 PDA 사용자들이 실시간 정보를 제공하는 웹 서버로 접속만 하면 업데이트 된 페이지를 바로 볼 수 있는 장점이 있다. 그러나 semi-automatic 트랜스코딩 방법이 지니는 단점으로 인하여 웹 페이지의 레이아웃이 변하게 되면 새로운 환경을 설정해야 하는 단점을 가지고 있다.

향후 연구과제로는 웹 페이지의 레이아웃이 변하더라도 실시간 속보 사이트의 URL이 변하지 않는 한 어떠한 경우에도 실시간 속보를 트랜스코딩 할 수 있는 방법에 대한 연구가 필요할 것이다.

## 참 고 문 헌

- [1] IBM, <http://www.kr.ibm.com/ibm/thinkers/fall2000/trend/index.html>.
- [2] J. R. Smith, R. Mohan and C. S. Li, "Transcoding Internet Content for Heterogeneous Client Devices", Proc. of ISCAS, June, 1998.
- [3] J. R. Smith, R. Mohan and C. S. Li, "Content-based Transcoding of Images in the Internet", Proc. of ICIP, Oct. 1998.
- [4] 안원섭, 나연목, "HTML 필터링을 통한 WML 컨텐츠 변환기의 설계 및 구현", 석사학위 논문, 단국대학교, 2000.
- [5] WAP Forum, <http://www.wapforum.org/>
- [6] IBM. WebSphere, <http://www.software.ibm.com/webservers/>
- [7] Hori M., Kondoh G., Ono K., Hirose S. and Singhal S., "Annotation-based Web Content Transcoding", Proc. of Ninth International WWW Conference, pp. 197~211, 2000.
- [8] Bickmore T. and Schilit W., "Digstor: device-independent access to the World Wide Web", Computer Networks and ISDN System, vol. 29, no. 8, pp. 1075~1082, 1997.
- [9] Brent A. Miller, Chatschik Bisdikian, Bluetooth

- Revealed, Prentice Hall, 2001.
- [10] Initium, <http://www.initium.co.kr>
- [11] Avantgo, <https://avantgo.com/frontdoor/index.html>
- [12] Ihandygo, <http://www.ihandygo.com/ihandygo/html/index.html>
- [13] Y. Whang, C. Jung, J. Kim, and S. K. Chung, "Webalchemist: A Web Transcoding System for Mobile Web Access in Handheld Devices", Proc. of SPIE, vol. 4534, Denver, Colorado, 2001.
- [14] Bickmore T., Girgensoh, A. and Sullivan J. W., "Web Page Filtering and Re-Authoring for Mobile Users", The Computer Journal, vol. 42, no. 6, pp. 534~546, 1999.
- [15] Pixa. Microbrowser 2.0, <http://www.pixo.com/products/products002.htm>

저 자 소 개



禹讚溢(正會員)

1993년 단국대학교 전자공학과 졸업(학사). 1995년 단국대학교 대학원 전자공학과 졸업(공학석사). 1997년~2000년 2월 단국대학교 대학원 전자공학과 박사과정 수료. 1995년~1997년 LG 이노텍 연구소. 2001년~현재 (주)비즈데이터. <주관심분야 : 정보보안, 전자상거래, 디지털 워터마킹, 멀티미디어>

1993년 단국대학교 전자공학과 졸업(학사). 1995년 단국대학교 대학원 전자공학과 졸업(공학석사). 1997년~2000년 2월 단국대학교 대학원 전자공학과 박사과정 수료. 1995년~1997년 LG 이노텍 연구소. 2001년~현재 (주)비즈데이터. <주관심분야 : 정보보안, 전자상거래, 디지털 워터마킹, 멀티미디어>



申仁澈(正會員)

1973년 고려대학교 전자공학과 졸업(학사). 1978년 고려대학교 대학원 전자공학과 졸업(공학석사). 1986년 고려대학교 대학원 전자공학과 졸업(공학박사). 1979~현재 단국대학교 전기전자컴퓨터공학부 교수. <주관심분야 : 병렬처리, 정보보안, 스마트카드, 전자상거래, 디지털 워터마킹>

1973년 고려대학교 전자공학과 졸업(학사). 1978년 고려대학교 대학원 전자공학과 졸업(공학석사). 1986년 고려대학교 대학원 전자공학과 졸업(공학박사). 1979~현재 단국대학교 전기전자컴퓨터공학부 교수. <주관심분야 : 병렬처리, 정보보안, 스마트카드, 전자상거래, 디지털 워터마킹>